A program to calculate the empirical bias in autocorrelation estimators

Jaume Arnau and Roser Bono
University of Barcelona

The statistical analysis of short time series designs is influenced by the presence of serial dependence. Therefore, it is important to correctly estimate the first-order autocorrelation in behavioral data. The empirical bias is an indicator generally used to evaluate the adequacy degree of an estimator. This paper presents the Bias program, a Monte Carlo simulation program that generates first-order autoregressive processes and calculates the bias in three autocorrelation estimators (r_1, r_1+ and r_1') for different values of the lag-one autocorrelation parameter and sample sizes. The program has been designed with MATLAB programming language and it runs on IBM-PC compatible computers with a 486 or later processor.

Programa para el cálculo del sesgo empírico de estimadores de autocorrelación. El análisis estadístico de los diseños de series temporales cortas está influido por la presencia de dependencia serial. De ahí la importancia de estimar correctamente la autocorrelación de primer orden en datos conductuales. El sesgo empírico es un indicador generalmente usado para evaluar el grado de precisión de un estimador. Este artículo presenta el programa Bias, un programa de simulación Monte Carlo que genera procesos autorregresivos de primer orden y calcula el sesgo de tres estimadores de autocorrelación (r_1, r_1+ y r_1') para diferentes valores del parámetro de autocorrelación de retardo uno y tamaños muestrales. El programa ha sido diseñado mediante el lenguaje de programación MATLAB y funciona en ordenadores IBM-PC con un procesador 486 o superior.

The study of autocorrelation in time series was considered from the end of the seventies (Bono & Arnau, 1996, 2000; Busk & Marascuilo, 1988; Escudero & Vallejo, 2000; Glass, Willson, & Gottman, 1975; Gorsuch, 1983; Hartman et al., 1980; Huitema, 1985, 1988; Jones, Vaught, & Weinrott, 1977; Rosel, Jara, & Oliver, 1999; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987; Vallejo, 1994, among others). In the nineties, the works were oriented towards the correction of the bias generated in short time-series by the conventional autocorrelation estimator r_1. Matyas and Greenwood (1991) verified that the difference between the mean of the conventional estimator r_1 and the value of the autocorrelation parameter ρ_1 increases with short time series. These authors concluded that the relationship between ρ_1 and the mean r_1 was nonlinear. The greatest coincidence between calculated mean r_1 and ρ_1 occurred when ρ_1 was around -0.2 or -0.3. According to these values, when ρ_1 increased positively, the calculated mean r_1 underestimated r_1 with a negative bias, while when the ρ_1 increased negatively, the calculated mean r_1 underestimated ρ_1, but with a positive bias. Based on this simulation study, it was deduced that when small sample sizes were used, the r_1 values were biased. Huitema and McKean (1991) proposed the modified estimator r_1+. This estimator corrects the empirical bias for positive values of the

autocorrelation parameter by adding $1/n$ to r_1. Arnau (1999) and Arnau and Bono (2001) proposed the estimator r_1' which consists of the correction of r_1 by the absolute value of a fitting model. This model is obtained from the polynomial function of the bias for each sample size. The estimator r_1' reduces the bias for small sample sizes, both for positive as well as negative values of ρ_1.

This article describes a computer program using MATLAB, Version 5.2 (1998). The program, called Bias, makes it possible to calculate the empirical bias in r_1, r_1+ and r_1' by Monte Carlo simulation. This program requires the specification of the parameter ρ_1 and the sample size (n). The structure of the program is as follows (see the Appendix): simulation of lag-one autoregressive processes, calculation of the estimators r_1 and r_1+, polynomial models for different sample sizes, calculation of the estimator r_1', and estimation of the empirical bias in r_1, r_1+ and r_1'. Bias runs under the Windows 3.1 or later operating system on IBM-PC compatibles with a 486 or later processor, and 8 MB of extended memory.

In the first place, the program simulates first-order autoregressive processes. The *randn* function (Forsythe, Malcolm, & Moler, 1977) is used to generate observations from a lag-one autoregressive process:

$$Y_t = \rho_1 Y_{t-1} + e_t \tag{1}$$

where Y_t is the score on the response measure at time t, ρ_1 is the autocorrelation parameter and e_t is a random normal variate with a mean of zero and a variance of one. Each time series starts with a normal variate (Y_0) having zero mean and variance $1/(1-\rho_1^2)$. For each combination of n and ρ_1, 10,000 samples are generated. The first 30 observations of each series are drop-

ped to eliminate any dependency between them (DeCarlo & Tryon, 1993; Huitema & Mckean, 1991, 1994a, 1994b, 1994c). After this, the program calculates the estimators r_1 (equation 2) and r_1+ (equation 3) for each simulated time series. The conventional estimator r_1 is defined by the following equation:

$$r_1 = \frac{\sum_{t=1}^{n-1}\left(Y_t - \overline{Y}_n\right)\left(Y_{t+1} - \overline{Y}_n\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}_n\right)^2} \qquad (2)$$

where Y_t and Y_{t+1} are the data obtained in the intervals t and $t+1$ and $\overline{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t$. The estimator r_1 is biased with small sample sizes because the numerator only includes n-1 terms instead of n. Huitema and McKean (1991) proposed the modified estimator r_1+ that corrects, to some extent, for the negative bias generated by positive autocorrelations

$$r_1+ = r_1 + (1/n) \qquad (3)$$

The empirical bias function of the conventional estimator r_1 mildly drifts from the linearity (Arnau, 1999; Arnau & Bono, 2001; DeCarlo & Tryon, 1993; Huitema & McKean, 1991, 1994a, 1994b). Models with significant coefficients of the empirical bias function for n= 6, 10, 20 and 30 were derived with polynomial fitting. Within the MATLAB environment, another program that calculated the empirical bias of the estimator r_1 in function of ρ_1 and n by Monte Carlo simulation was designed (Arnau, 1999; Arnau & Bono, 2001). The *polytool* function of the *Statistics Toolbox* of the MATLAB was used for polynomial fitting. With the significant polynomial coefficients at 5% (Table 1) the following fitting models based on n were designed and were introduced into the Bias program.

$Model_{n=6}$= -0.1648 – 0.5643r_1 – 0.0916r_1^2 $\qquad (4)$
$Model_{n=10}$= -0.0972 – 0.3760r_1 – 0.0676r_1^2 $\qquad (5)$
$Model_{n=20}$= -0.0482 – 0.2028r_1 – 0.0333r_1^2 $\qquad (6)$
$Model_{n=30}$= -0.0373 – 0.1360r_1 $\qquad (7)$

For sample sizes other than those studied in this paper, it is previously necessary to calculate, by simulation, the empirical biases of the conventional estimator. Once the corresponding empirical biases have been found, the polynomial fitting is carried out with *polytool* function specifying the ρ_1 values, empirical biases previously obtained and the degree of the polynomial fit. In this way, the significant coefficients are obtained. These must be included in the *polynomial models for different sample sizes* statement indicating the value of n.

Using the equations 4-7, the Bias program calculates the estimator r_1' by adding the corresponding polynomial model in absolute values to r_1.

$$r_1' = r_1 + Fitting\ Model$$
$$(8)$$

Finally, for each simulated sample, the program calculates the empirical bias in r_1, r_1+ and r_1', which is the difference between the mean value of the estimator and the value of the parameter ρ_1.

The Bias program only calculates r_1' for four sample sizes (n= 6, 10, 20 and 30). For other n values, the polynomial fitting of the bias must be performed first in order to determine the corresponding correction model and to introduce it into the program.

Running the program

The Bias program is a function M-file and the input variables in the MATLAB workspace, autoregressive parameter (rho) and sample size (n), must be introduced into it. As can be seen in the Appendix, the first line of the function M-file specifies the file name (bias), the input variables names (rho and n) and output variables names (r1_bias, r1plus_bias and r1prime_bias). The following expression must be specified in the MATLAB workspace in order to run the program: [r1_bias, r1plus_bias, r1prime_bias]= bias(rho,n). The MATLAB executes the commands in Bias program and the results are shown in the MATLAB workspace.

Availability

Interested users who prefer not to type the program can request a file of the program by e-mailing the second author (rbono@psi.ub.es). It has been designed with the Version 5.2 of MATLAB but it can also work with previous versions.

Acknowledgements

Table 1 Coefficients of the polynomic functions of the bias						
		Model				
n		**Constant**		**Linear**		**Quadratic**
6	Coefficients:	-0.1648		-0.5643		-0.0916
	CI (0.05)	-0.1703	-0.1592	-0.5725	-0.5561	-0.1260 -0.0572
10	Coefficients:	-0.0972		-0.3760		-0.0676
	CI (0.05)	-0.1039	-0.0904	-0.3860	-0.3660	-0.1096 -0.0257
20	Coefficients:	-0.0482		-0.2028		-0.0333
	CI (0.05)	-0.0530	-0.0433	-0.2100	-0.1957	-0.0634 -0.0032
30	Coefficients:	-0.0373		-0.1360		
	CI (0.05)	-0.0406	-0.0341	-0.1470	-0.1251	
CI = Confidence Interval						

APPENDIX
Program Listing

```
function [r1_bias, r1_plus_bias, r1_prime_bias]=bias(rho,n)

%
***************************************************
% LAG-ONE AUTOREGRESSIVE PROCESSES SIMULA-
TION
%
***************************************************
% Number of simulations
for s=1:10000
% Number of observations
    obs=30+n;
% Variance
    variance=1/(1-rho^2);
% Standard deviation
    sd=sqrt(variance);
% Series with mean 0 and fixed standard deviation
    y0=normrnd(0,sd,obs,1);
% Random series with mean 0 and variance 1
    e=randn(obs,1);
% First value of the series
    y(1,1)=y0(1)+e(1);
% Next values of the series
    for t=2:obs
        y(t,1)=rho*y(t-1,1)+e(t);
        t=t+1;

end
% Eliminations of the first 30 observations
y=y(31:obs);


% ***************************************************
% CALCULATION OF r1
% ***************************************************
y_mean=mean(y);
denominator=(y-y_mean).^2;
denominator_sum=sum(denominator);
y_lag=y(2:n);
y=y(1:n-1);

y_lag_diff=y_lag-y_mean;
y_diff=y-y_mean;
numerator=y_lag_diff.*y_diff;
numerator_sum=sum(numerator);
r1(s,1)=numerator_sum/denominator_sum;
end


% ***************************************************
% CALCULATION OF r1+
% ***************************************************
r1_plus=r1+(1/n);


%
***************************************************
% POLYNOMIAL MODELS FOR DIFFERENT SAMPLE SI-
ZES
%
***************************************************
if n==6
    model = -[(-0.1648)+r1.*-0.5643+(r1.^2).*-0.0916];
elseif n==10
    model = -[(-0.0972)+r1.*-0.3760+(r1.^2).*-0.0676];
elseif n==20
    model = -[(-0.0482)+r1.*-0.2028+(r1.^2).*-0.0333];
elseif n==30
    model = -[(-0.0373)+r1.*-0.1360];
end


% ***************************************************
% CALCULATION OF r1'
% ***************************************************
r1_prime=r1+model;


% ***************************************************
% EMPIRICAL BIAS IN r1, r1+ AND r1'
% ***************************************************
r1_bias=mean(r1)-rho;
r1_plus_bias=mean(r1_plus)-rho;
r1_prime_bias=mean(r1_prime)-rho;
```

References

Arnau, J. (1999). Bias reduction in the estimation of autocorrelation in short time series. *Metodología de las Ciencias del Comportamiento, 1*, 25-37.

Arnau, J. & Bono, R. (2001). Autocorrelation and bias in short time series: an alternative estimator. *Quality and Quantity, 35*, 365-387.

Bono, R. & Arnau, J. (1996). *C* statistic power analysis through simulation. *Psicothema, 8*, 699-708.

Bono, R. & Arnau, J. (2000). Designs of small samples: Analysis by generalized least squares. *Psicothema, 12 Suppl. 2*, 87-90.

Busk, P.L. & Marascuilo, L.A. (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment 10*, 229-242.

DeCarlo, L.T. & Tryon, W.W. (1993). Estimating and testing autocorrelation with small samples: a comparison of the *C*-statistic to a modified estimator. *Behavior Research Theory, 31*, 781-788.

Escudero, J.R. & Vallejo, G. (2000). Comparison of three alternative methods for the analysis of interrupted time-series. *Psicothema, 12*, 480-486.

Forsythe, G.E., Malcolm, M.A. & Moler, C.B. (1977). *Computer methods for mathematical computations*. Englewood Cliffs, NJ: Prentice Hall.

Glass, G.V., Willson, V.L. & Gottman, J.M. (1975). *Design and analysis of time series experiments*. Boulder: Colorado Associated University Press.

Gorsuch, R.L. (1983). Three Methods for analyzing limited time-series (*N* of 1) data. *Behavioral Assessment, 5*, 141-154.

Hartmann, D.P., Gottman, J.M., Jones, R.R., Gardner, W., Kazdin, A.E. & Vaught, R.S. (1980). Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis 13*, 543-559.

Huitema, B.E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment 7*, 107-118.

Huitema, B.E. (1988). Autocorrelation: 10 years of confusion. *Behavioral Assessment 10*, 253-294.

Huitema, B.E. & McKean, J.W. (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin, 110*, 291-304.

Huitema, B.E. & McKean, J.W. (1994a). Two reduced-bias autocorrelation estimators: r_{F1} and r_{F2}. *Perceptual and Motor Skills, 78*, 323-330.

Huitema, B.E. & McKean, J.W. (1994b). Reduced bias autocorrelation estimation: Three jackknife methods. *Educational and Psychological Measurement, 54*, 654-665.

Huitema, B.E. & McKean, J.W. (1994c). Test of H_0: $\rho_1 = 0$ for autocorrelation estimators: r_{F1} and r_{F2}. *Perceptual and Motor Skills, 78*, 331-336.

Jones, R.R., Vaught, R.S. & Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis 10*, 151-166.

MATLAB. (1998). MATLAB (Version 5.2). Natick, MA: The MathWorks, Inc.

Matyas, T.A. & Greenwood, K.M. (1991). Problems in the estimation of autocorrelation in brief time series and some implications for behavioral data. *Behavioral Assessment, 13*, 137-157.

Rosel, J., Jara, P. & Oliver (1999). Cointegration in multivariate time series. *Psicothema, 11*, 409-419.

Sharpley, C.F. & Alavosius, M.P. (1988). Autocorrelation in behavioral data: An alternative perspective. *Behavioral Assessment 10*, 243-251.

Suen, H.K. (1987). On the epistemology of autocorrelation in applied behavior analysis. *Behavioral Assessment 9*, 113-124.

Suen, H.K. & Ary, D. (1987). Autocorrelation in applied behavior analysis: Mith or reality?. *Behavioral Assessment 9*, 125-130.

Vallejo, G. (1994). Evaluation of the effects of intervention in designs of time series in the presence of trends. *Psicothema, 6*, 503-524.